

# Invenio@HGF – status and perspectives

[sic!] Jülich – 2nd Invenio User Group Workshop

Jülich, 18. November 2013 | Alexander Wagner, for the Collaboration

## Overview



- Partner
- Initial TODO
- Accomplishments
- Lessons learned
- Project group

## Project Partners



Deutsches Elektronensynchrotron, Zentralbibliothek

≈ 2000 + 3000



Forschungszentrum Jülich, Zentralbibliothek

≈ 5000 + 1000



GSI Helmholtzzentrum für Schwerionenforschung, Bibliothek + Kern-IT

≈ 1050



Maier-Leibniz-Zentrum, Garching

≈ 300



RWTH Aachen, Hochschulbibliothek

≈ 9000

Museum Zitadelle Jülich

Institut für Experimentelle Kernphysik, Karlsruhe

### Open for new Partners!

Serving now ≈ **17.000** people (+ visitors)

(≈ 260.000 documents + 67.000 Authorities)



Alexander Wagner, for the Collaboration

3 | 16

## Starting out

### Goal

Replace existing systems, **at GSI** build up from scratch.  
User-centric design (users aka **scientists**)

- 1 "Learn Invenio" (thanks to CERN ☺)
- 2 Define wording. . . (different institutions!)
- 3 Build infrastructure: **git** and friends
- 4 Build more infrastructure: **authorities** and friends
- 5 Build a **deployment scheme**: **InstallInvenio** and friends

We need to roll out **10+** instances  
with **different** data sets and **keep them consistent** on code level

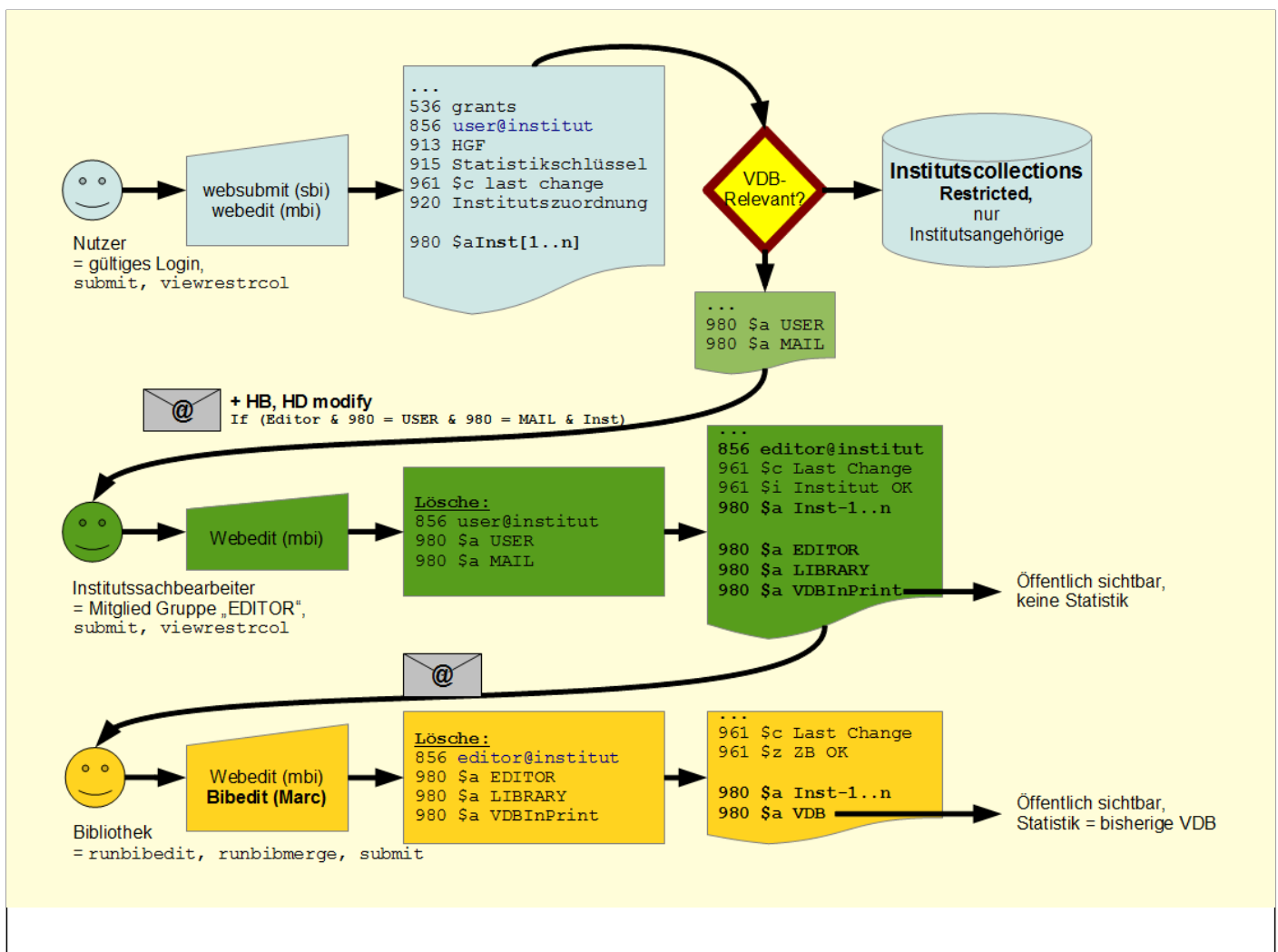


Alexander Wagner, for the Collaboration

4 | 16

## Initial ToDo

- Design the system around **web based literature management**
- Design a document workflow



## Initial ToDo

- Design the system around **web based literature management**
- Design document workflow (3 steps with privilege escalation)
- Design easy ingestion workflow (websubmit, imports, author disambiguation)

Finally we wrote some code. . .

Every unwritten line is a good line, still:  $\approx 55.000$  lines

- Migrate old data (various, proprietary sources)
- Train the inputters and users (secretaries, scientists, librarians)
- Hook up with content management system(s) (visibility!)



Alexander Wagner, for the Collaboration

7 | 16



The screenshot shows a web browser window displaying the Jülich Forschungszentrum website. The page is titled "Peter Grünberg Institut (PGI) Quanten-Theorie der Materialien (PGI-1 / IAS-1)". The main navigation bar includes links for "AKTUELLES", "FORSCHUNG", "LEISTUNGEN", "KARRIERE", and "ÜBER UNS". The "Publikationen" section is active, showing a list of publications for 2013. The list includes titles such as "Slater-Pauling behavior in LiMgPdSn-type multifunctional quaternary Heusler materials", "GW study of topological insulators Bi<sub>2</sub>Se<sub>3</sub>, Bi<sub>2</sub>Te<sub>3</sub>, and Sb<sub>2</sub>Te<sub>3</sub>", and "Precise response functions in all-electron methods: Generalization to nonspherical perturbations and application to NiO". The page also features a sidebar with links to "Kontakt und Anfahrt", "Mitarbeiter", and "Publikationen".

## Initial ToDo

- Design a document workflow (3 steps with privilege escalation)
- Establish easy ingestion workflow (websubmit, imports, author disambiguation)

Finally we wrote some code. . .

Every unwritten line is a good line, still:  $\approx 55.000$  lines

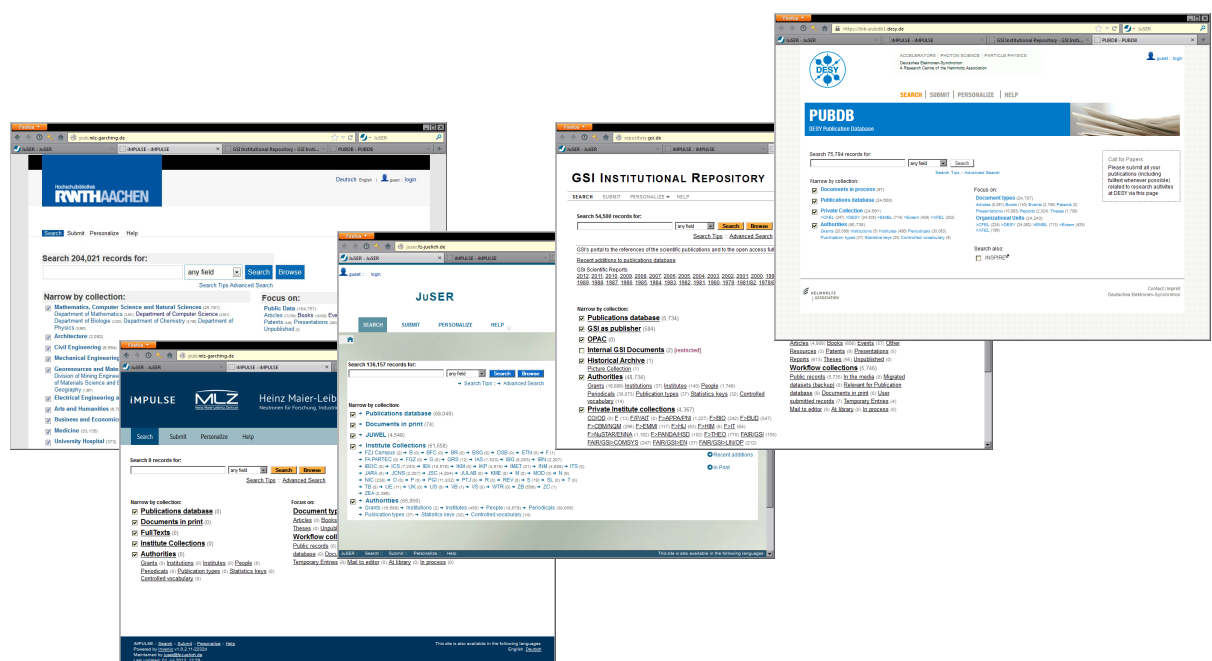
- Migrate old data (various, proprietary sources)
- Train the inputters and users (secretaries, scientists, librarians)
- Hook up with content management system(s) (visibility!)
- Derive necessary reporting (details are way beyond this intro)
- Get it up and running (First Light: 11/19/2012)



Alexander Wagner, for the Collaboration

9 | 16

## Accomplishments and status



Alexander Wagner, for the Collaboration

10 | 16

## Accomplishments and status

- All partners have running systems (roll out works)
- Almost all partners are online
- Rich websubmit (including **repeatable field** handling)
- Importer routines (doi, pmid, arXiv, inspire, ISBN, own recs, ... in **websubmit**)
- **Authorities**
  - **Generate** ( $\approx 67.000$  recs)
  - **Use** (e. g. JSON returns, statistics. ...)
  - **Share** (MarcXML OAI-PMH)
- **Implement**
  - **Author identification** (ORCID ready!)
  - **Output formats** (JSON, BibT<sub>E</sub>X, EndNote. ...)
  - **Reporting** (publication statistics)
  - **Delivery to content management systems**



## Tools used

- **Workflow**
  - **Webbaskets** (e. g. revision lists)
  - **Alerts** (e. g. revision lists)
  - **Collections** (e. g. private for institutes)
  - **Webmessage** (e. g. correction requests)
- **Authority records** (almost everywhere)
- **OAI-PMH** (authority exchange)
- **High-level API** (setup: e. g. collections, roles, groups, baskets. ... ; **no db-dump** sharing)
- **jQuery/jQueryUI** (websubmit)
- **intbitsets** (e. g. statistics)



## Lessons learned / Next steps

- CERN is *way to fast* to keep up with
- Never use Dublin Core again (*complex migration, to few data fields...*)
- All libraries are the same ☺
- *Upgrade to 1.2*: get OAI-Server fixed!

### However...

In our use case switching of the base system is non-trivial

(Remember: 10+ instances...)

- Open up for new partners
- Clean up our code and give it back



Alexander Wagner, for the Collaboration

13 | 16

## Contributors

- |   |  |
|---|--|
| ■ <i>Martin Köhler<sup>a</sup></i>        | ■ <i>Alexander Wagner<sup>e</sup></i>  |
| ■ Zaven Akopov <sup>a,b</sup>             | ■ <i>Jürgen Neuhaus<sup>f</sup></i>    |
| ■ <i>Tomasz Pazera<sup>a</sup></i>        | ■ <i>Connie Hesse<sup>f</sup></i>      |
| ■ <i>Katrin Große<sup>c</sup></i>         | ■ <i>Björn Pedersen<sup>f</sup></i>    |
| ■ <i>Stefan Hesselbach<sup>d</sup></i>    | ■ <i>Ulrike Eich<sup>g</sup></i>       |
| ■ <i>Bernhard Mittermaier<sup>e</sup></i> | ■ <i>Louai Barake<sup>g</sup></i>      |
| ■ <i>Anna Fründ<sup>e</sup></i>           | ■ <i>Abdoulaye Diallo<sup>g</sup></i>  |
| ■ <i>Heike Lexis<sup>e</sup></i>          | ■ <i>Roland Rappmann<sup>g</sup></i>   |
| ■ <i>Cornelia Plott<sup>e</sup></i>       | ■ <i>Dominik Schmitz<sup>g</sup></i>   |
| ■ <i>Christoph Holzke<sup>e</sup></i>     | ■ <i>Edmund Wollgarten<sup>g</sup></i> |

<sup>a</sup> DESY Library and Documentation; <sup>b</sup> Project Inspire; <sup>c</sup> GSI Library; <sup>d</sup> GSI Core IT;

<sup>e</sup> Forschungszentrum Jülich, Zentralbibliothek; <sup>f</sup> MLZ, Garching; <sup>g</sup> RWTH Aachen, Hochschulbibliothek



Alexander Wagner, for the Collaboration

14 | 16

## Further “reading”

- Invenio @ HGF - Technical background

Talk at Invenio Developer Forum

- Collaborative tools for an institutional repository

Talk at Helmholtz OA Workshop

- JuSER – Publications Database

Introductory course at Jülich

- JuSER - Autorenhandling

Talk at HGF-ORCiD Meeting, Berlin (in german)



Alexander Wagner, for the Collaboration

15 | 16

## Thanks!



Alexander Wagner  
Zentralbibliothek

Scientific Services /  
Scientific Publishing

Tel.: +49-2461-61-1586  
a.wagner@fz-juelich.de

This document is available as FZJ-2013-05410



Typeset by pdfL<sup>A</sup>T<sub>E</sub>X



Alexander Wagner, for the Collaboration

16 | 16